**Fundamentals of Multimedia Computing**

**Chapter 5: More Sensors**

**Authors:**

**Gerald Friedland**

**and**

**Ramesh Jain**

**Draft for Comments**

**More Sensors and Multimodal Integration**

The sensors for sound and light described in the previous two chapters have been the most important for multimedia computing in the last decades because audio and video are best for communicating information for the tasks typically performed with a computer. That is, most people prefer to communicate through sound, and light serves illustrative purposes, supplementing the need for language-based description of a state of the world. New or different tasks might use different sensors, however. For example, in dating, communication probably occurs on many other levels, such as scent, touch, and taste **//Are you referring to online dating? If so, “communication would probably occur” would be more correct because current sites don’t offer the other types of sensors.//**. In this chapter we discuss other sensors, including sensors that are not part of the human sensory system, and then touch on multimodal integration—the integration of different sensors into one integrated sensation.

**Properties and Types of Sensors**

In general, a sensor is a device that measures a physical quantity and converts it into a signal that an observer or instrument can read. Whether the sensor is human-made or from nature does not matter. An ideal sensor is sensitive to the measured property, insensitive to any other property, and does not influence the measured property. Of course, no perfect sensor exists because the laws of physics state that energy is conserved and sensors need a transfer of energy to function. After all, the **photos //photons?//** absorbed in the retina of a human eye interfere with the universe in the sense that **//if they were not absorbed by the human eye,//** they would not be absorbed at all or would be absorbed by a different object. In practice, however, these rules help optimize sensor development. They also help to distinguish common deviations from ideal sensors; in fact, it’s these deviations that multimedia computing uses for compression, corrects when reproducing signals, and analyzes for content.

If the sensor is not ideal, several types of deviations can typically be observed.

First, no sensor has unlimited range. Therefore, the sensor might saturate; that is, whereas the ideal measurement response would suggest a further increase in output, the sensor outputs a maximum saturated value and/or breaks (compare, for example, human ears exposed to too-loud noises). The range’s lower bound is defined by the minimum amount of input that can be clearly distinguished from no input. If the output is not zero when the input is zero, the sensor is said to have an *offset* or *bias*. In practice, the sensitivity might differ from the measurement function specified. Ideally, a sensor will respond linearly to the measured entity; that is, the sensitivity should be constant. This error is therefore often described as *nonlinear behavior*.

Sensors are often tuned to behave linearly inside an operational range. The error might be dynamic in that the sensor might behave differently based on time or other influencing factors that vary independently from the measured entity. A changing sensitivity given a constant signal is a *drift*. Most sensors have long-term drift due to aging. A random deviation from the measurement function is called *noise*.

*Hysteresis* is a deviation over time: When the measured entity reverses direction (for example, gets higher instead of lower) but the sensor’s response has a finite lag, it might create one offset in one direction and a different offset in the other. **Figure 1** **//renumber so Figures start at #1 in chapter one and continue rather than restart?//** illustrates the concept. Errors resulting from sampling are called *digitization errors* and *aliasing errors*; see Chapter XXX for an in-depth discussion. A sensor’s resolution is the smallest change it can detect in the quantity that it is measuring. Resolution might also behave nonlinearly.

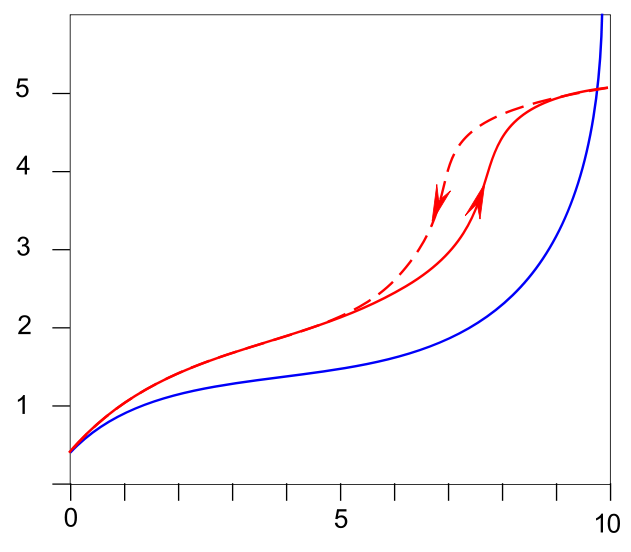


Figure 1. A general example for the hysteresis concept. The bottom curve (blue) is the sensor input, and the upper curve is the output. As the arrows show, the curve behaves differently when the measured entity decreases compared to when it increases.

In addition to light and sound, many animals have sensors for temperature, gravity, humidity, vibration, pressure, smell, and other properties of their environment. Many nerve systems and chemical sensors also sense internal aspects of the body. Some artificial sensors can sense physical as well as chemical phenomena, such as radiation, force, flow, and seismic activity. Most of these sensors are not (yet?) of interest in multimedia computing.

Touch is seeing increased use in multimedia computing. Haptic technologies use tactile sensors to translate human motion into computation. Currently, the most common type of tactile sensor is force feedback, which is used in applications such as computer game controllers, servo mechanisms for aircrafts, and remote surgery. Haptic devices receive information from the user through pressure-sensitive sensors, such as piezo crystals, which emit electrical power proportional to the force applied to them. Actuators transmit the feedback using electro motors to induce a vibration or another physical motion to communicate to the human hands and arms. Other sensors, such as haptic gloves, allow normal usage of the hand but capture the muscular state of the fingers and other parts of the hand. Often, a remote robotic hand then reproduces the state of the operator’s hand on the other end. A major issue with haptic sensors is tuning the feedback or sensor to make the user experience intuitive and natural, so the user does not have to learn how to operate the system. Although this is easily achieved with conventional video and audio, haptic sensors often require combining different sensors and feedback—which leads us to the topic of multimodal integration.

**Multimodal Integration**

The human brain can integrate different sensory modalities, such as sight, sound, and touch, into a coherent and unified perceptual experience. Experiments show that by considering input from multiple sensors, perceptual problems can be solved more robustly and even faster (see references). This *multimodal integration*, or *multisensory integration*, is not yet completely understood, but it is fundamental to the success of multimedia systems. Multimedia computing strives to imitate the properties of multimodal integration regardless of the incomplete understanding of the mechanisms in the brain. or example, multimedia content analysis (as described in Chapter XXX) combines audio and video in an attempt to gain accuracy, robustness, and sometimes speed.

Here, we describe some well-known observable phenomena that might help to both clarify the process and highlight the design considerations for multimedia systems.

Experiments have indicated that two converging sensory stimuli can produce a perception that differs not only in magnitude from the sum of the two individual stimuli but also in quality. The classic study, which introduced the *McGurk effect*, dubbed a person’s acoustic phoneme production with a video of that person speaking a different phoneme. The result was that the user perceived neither the visual nor the acoustic pronunciation but instead heard a third phoneme. McGurk and MacDonald explained in their 1976 article that phonemes such as ba, da, ka, ta, ga, and pa can be divided into two groups:

* phonemes that can be visually confused (da, ga, ka, ta), and
* phonemes that can be acoustically confused (ba and pa).

The combination of the visual and acoustically confused phonemes results in the perception of a different phoneme. For example, when an uttering of ba is dubbed on a video showing the uttering of ga, which is processed visually through lip reading, the visual modality sees ga or da, and the auditory modality hears ba or da, which combine to form the perception of da.

Ventriloquism is another important effect. It describes the situation in which acoustic tracking of a sound’s origin shifts toward the visual modality. In conditions where the visual cue is unambiguous, the perception of the visual location overrides the acoustic location. Artists throughout the world use this effect. Ventriloquists manipulate how they produce sound so it appears that the voice is coming from elsewhere, usually a puppet.

An almost “magic” effect is called *body transfer illusion*. Botvinick and Cohen performed the original, so-called rubber hand experiment in 1998. Human participants sat in front of a screen showing a dummy hand being stroked with a brush while they felt a series of synchronized and identical brushstrokes applied to their own hands, hidden from their view. The result was that if the dummy hand was similar to the participant’s hand in appearance, position, and location, the human subject was likely to feel that the touches on his or her hand came from the dummy hand **//correct?//**. Furthermore, several participants reported that they felt the dummy hand to be their own hand. Virtual reality applications exploit this effect and try to induce the perception of owning and controlling someone else’s body (usually an avatar) by applying visual, haptic, and sometimes proprioceptual stimulation synchronously.

The brain exploits multimodal integration in different ways. The two most important are probably the decrease of sensory uncertainty and the decrease of reaction time. Experiments have shown that uncertainty in sensory domains leads to an increased dependence of multisensory integration. If a person sees something moving in a tree and isn’t sure whether it is a bird or a squirrel, the natural reaction is to listen. If the object emits a chirp and the brain localizes the sound to be coming from the tree, the person takes this as proof that the creature is a bird. Hence, the lack of visual information is augmented by acoustic information.

The Hershenson (YEAR) experiments also showed that responses to multiple simultaneous sensory stimuli can be performed faster than responses to the same stimuli presented in isolation. Participants were presented a light and tone simultaneously and separately, and were asked to respond as rapidly as possible by pressing a button. Reaction time differed with varying levels of synchrony between the tone and the light. **The former result** **//I’m not sure which result this is.//** is, however, hard to generalize as multiple synchronous stimuli might also cause the opposite effect, as we discuss in the next section.

**Split Attention**

*Split attention*, the opposite effect of multimodal integration, manifests when the same media (for example, visual and visual) is used for different types of information at the same time. To understand and use the materials provided, one must split attention between them. Split attention should not be confused with distraction, although the two problems are related. Distraction is caused by the lack of ability to pay attention to a particular object due to lack of interest in the object or the greater intensity, novelty, or attractiveness of something other than the object of attention. However, split attention is caused by the lack of integration of the object to be paid attention to.

Figure 2 shows an example multimedia system known to have caused a split attention problem. The E-Chalk lecture recording system showed dynamic board content in addition to a video of the lecturer. In a typical E-Chalk lecture, two areas of the screen compete for the viewer’s attention: the video window showing the instructor and the board or slides window. Several researchers tracked students’ eye movements as they watched a lecture recording containing slides and an instructor video in a setup similar to that shown in Figure 2 (see references). Measurements showed that a student often spends about 70 percent of the time watching the instructor video and only about 20 percent of the time watching the slides. For the remaining 10 percent of time, the eye focus was lost to activities unrelated to lecture content. When the lecture replay consists of only slides and audio, students spend about 60 percent of the time looking at the slides because they have no other place to focus their attention in the lecture recording. The remaining 40 percent, however, was lost due to distraction.

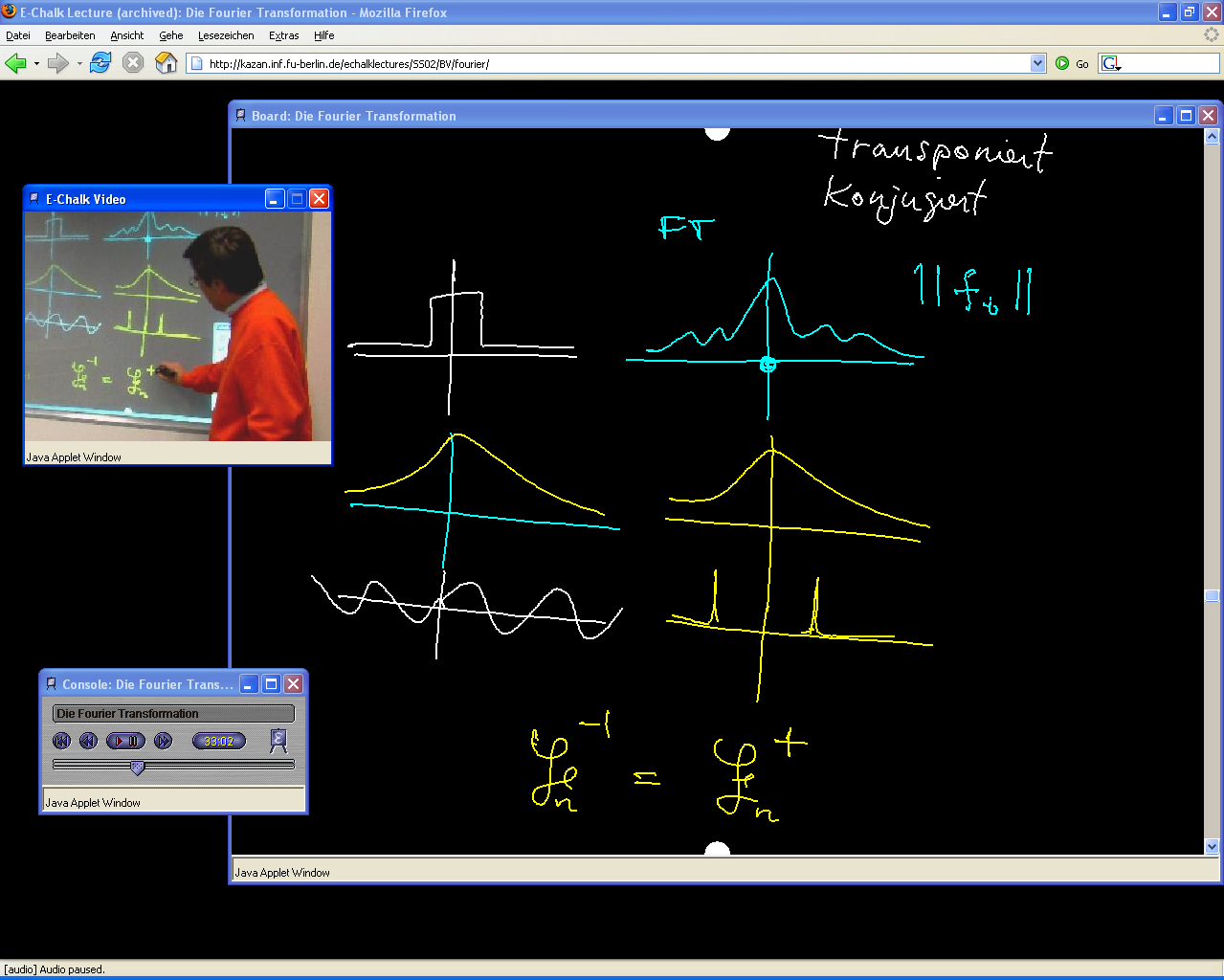


Figure 2. An example of a split-attention problem. The lecturer on the left is shown as a video, and the dynamic board content he creates is shown on the right. The additional video window is used to convey gestures and finger pointing. However, presenting the lecturer in a second window causes cognitive overhead usually referred to as split attention because the viewer switches between the content on the board and the lecturer on the left.

It is still an open question whether attention can be split to attend to two or more sources of different (???) information simultaneously. Psychologists and neuroscientists have discussed this topic for decades. Most researchers, however, now accept that attention can be split at the cost of cognitive overhead. It is this cognitive overhead that is to be avoided when designing multimedia systems. See the references for more information **//Will references be set for each chapter, or only at the end, with additional readings for each chapter?//**.

So far, we have described sensory phenomena on a high level and discussed important properties of the human system. We will now begin to dig deeper into what it means to process sensory information using computers. While reading the next chapters you might find it useful to go back to these introductory chapters from time to time to remind yourself of the fundamentals. However, you will do so at the cost of some cognitive overhead, a phenomenon called split attention.

**//Include key terms and/or key points for each chapter?//**

**Literature**

* C. A. Grimes, E. C. Dickey, and M. V. Pishko (2006), Encyclopedia of Sensors (10-Volume Set), American Scientific Publishers.
* Gemma Calvert, Charles Spence, Barry E. Stein: The handbook of multisensory processes, MIT Press,2004.
* Baars, B. J. (1988): A Cognitive Theory of Consciousness. Cambridge University Press, Cambridge, UK, 1988.

**Research Articles**

* A. Riul, Jr., R. R. Malmegrim, F. J. Fonseca, and L. H. C. Mattos: An artificial taste sensor based on conducting polymers, Biosensors and Bioelectronics, Volume 18, Issue 11, October 2003, Pages 1365-1369.
* Krishna Persaud, George Dodd: Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose, Nature 299, 352 - 355 (23 September 1982).
* M. Hershenson: Reaction time as a measure of intersensory facilitation. J Exp Psychol, 63:289–93, 1962.
* H. McGurk and J. MacDonald: Hearing lips and seeing voices. Nature, 264(5588):746–48, 1976.
* Hahn, S. and Kramer, A. F. (1998). Further evidence for the division of attention among non-contiguous locations. Visual Cognition, 5(1-2):217–256.
* Narcisse P. Bichot and Kyle R. Cave and Harold Pashler (1999). Visual selection mediated by location: Feature-based selection of non-contiguous locations. Perception & Psychophysics, 61(3):403–423.
* Chandler, P. and Sweller., J. (1992). The Split Attention Effect as a Factor in the Design of Instruction. British Journal of Education Psychology, 62:233–246.
* Sweller, J., Chandler, P., Tierney, P., and Cooper, G. (1990). Cognitive Load as a Factor in the Structuring of Technical Material. Journal of Experimental Psychology: General, 119:176–192.
* Glowalla, U. (2004). Utility und Usability von E-Learning am Beispiel von Lecture-on-demand Anwendungen. Fortschritt-Berichte VDI, 22(16):603–621.
* Mertens, R., Friedland, G., and Krueger, M. (2006). To See or Not To See: Layout Constraints, the Split Attention Problem and their Implications for the Design of Web Lecture Interfaces. In Proceedings of the AACE E-Learn – World Conference on E-Learning in Corporate, Govern- ment, Healthcare, and Higher Education, Honolulu, Hawai, USA.
* Botvinick, M. & Cohen, J. D. (1998). Rubber hand ‘feels’ what eyes see. Nature, 391, 756.

**Exercises**

1. Traditional wisdom gives humans five senses: vision, hearing, touch, smell, and taste. Explain why this number is wrong and discuss why it is not easy to define what a sense is.
2. Habituation describes an effect in which repeated exposure to a stimulus leads to lower response. Provide examples of human sensors that show habituation.
3. Explain why haptic sensors in virtual reality require physical robustness beyond what is often technically achievable.
4. Explain why the McGurk effect is generally not a problem for dubbing TV shows and movies into a different language.
5. Find at least one more example for the split attention issue in daily life. Propose a solution.